

Statistical learning for resting-state fMRI: successes and challenges

G. Varoquaux¹²³, B. Thirion¹²

¹ INRIA, Parietal team, Saclay, France

² LNAO/NeuroSpin, CEA Saclay, Bat. 145, 91191 Gif-sur-Yvette, France

³ INSERM U992, NeuroSpin center, 91191 Gif-sur-Yvette, France

Abstract. In the absence of external stimuli, fluctuations in cerebral activity can be used to reveal intrinsic structures. Well-conditioned probabilistic models of this so-called resting-state activity are needed to support neuroscientific hypotheses. Exploring two specific descriptions of resting-state fMRI, namely spatial analysis and connectivity graphs, we discuss the progress brought by statistical learning techniques, but also the neuroscientific picture that they paint, and possible modeling pitfalls.

1 Introduction

The study of the fluctuations in on-going brain activity, present at rest, opens a window on intrinsic functional brain structures. Indeed, correlation patterns observed in fMRI without external stimuli mirror, to some extent, task-related networks [4]. Using the resting brain to accumulate knowledge on its function requires statistical models of the signal, though too often descriptive statistics, such as sample correlations, are used to draw inferences. The estimation of such models is limited, at the subject level, by the scarcity of the data available, and, at the population level, by intersubject variability.

Here we discuss what, from our point of view, statistical machine learning has brought to resting-state modeling, namely well-conditioned models and estimators in high-dimensional settings due to a mature approach to regularization. Learning higher-dimensional models brings richer neuroscientific pictures, but all the modeling issues cannot be addressed with machine learning. Due to its brevity, this paper presents only a small view of the resting-state modeling literature. First we revisit the spatial extraction of resting-state networks and show that adding penalization to models yields finer-grained descriptions. Second we consider inferring connectivity graphs, for which regularization is crucial but can also shape our understanding of brain wiring.

2 Spatial analysis: better models define finer networks

2.1 A handful of large-scale brain networks or a score of regions?

Early insight on resting-state came from seed-analysis studies, such as in the motor cortex [4], that highlighted brain networks correlating with a given seed

region during rest. However, a wide range of different networks can be obtained by varying the choice of the seed [8]. For this reason, investigators have turned to multivariate methods that do not require such arbitrary choice, in particular Independent Component Analysis (ICA), to extract the spatial modes of resting-state activity. These approaches have identified a small number of large-scale brain networks consistently found across studies and subjects [9]. Recent studies argued that these networks could be divided in a set of smaller consistent brain regions [18, 13]. From a multivariate analysis point of view, the complexity of the models that can be learned is limited by the amount of data at hand. Is the small number of networks an over simplification due to data scarcity? At which scale can one extract brain networks?

2.2 The importance of penalizing: ICA to sparse decompositions

While data-driven methods such as ICA are often called *model-free*, they rely on the following simple *multivariate decomposition* model of the signal: $\mathbf{Y} = \mathbf{U}\mathbf{V} + \mathbf{N}$, where $\mathbf{Y} \in \mathbb{R}^{n \times p}$ are the fMRI data: n observations of images with p voxels, $\mathbf{U} \in \mathbb{R}^{n \times k}$ and $\mathbf{V} \in \mathbb{R}^{k \times p}$ are respectively k time series and k spatial maps learned from the data, and $\mathbf{N} \in \mathbb{R}^{n \times p}$ some additional noise. This model needs additional constraints to be well-posed. ICA imposes low mutual-information between the spatial maps. However, it has no intrinsic noise model: it fits all the available signal. Thus, in practice, ICA is most often combined with an preliminary PCA step to select a smaller signal subspace accounting for most of the original signal variance [1]. In addition, output maps are thresholded to remove background noise. The resulting maps that are considered as neuroscientifically meaningful are spatially structured and highlight a small number of activated brain regions. ICA also segments artifacts in the fMRI data, such as vascular or movement related signal. However, if the model order, *i.e.* the number of components, is increased, some output maps show little structure and are difficult to interpret. This limitation stems from the inability of ICA to reject noise: at high model order the PCA step used for this purpose includes signal that cannot be well decomposed according to the ICA criterion. The two step estimation PCA + ICA is necessary as the ICA criterion cannot be combined with a criterion controlling explained variance. In Bayesian terms, it corresponds to an improper prior. A solution to this limitation is to frame the model in terms of sparsity, rather than independence. Indeed ICA selects sparse maps as a byproduct of the independence criterion, and this sparsity is central to its success on fMRI data [10, 17].

An ℓ_1 penalization can impose a sparse prior on the spatial maps \mathbf{V} in the decomposition model. Combined with PCA in a *sparse PCA* procedure, it regularizes model estimation by constraining the fitted signal [19]. As ℓ_1 penalizations perform poorly with correlated features, it is useful to impose sparsity on groups of neighboring voxels, for instance using sparse structured penalties. A *Sparse structured PCA* (SSPCA) [12] extracts the regions comprising the brain networks in separate maps (see Fig. 1). At high model order SSPCA remains stable and interpretable as it learns finer-grained segmentations of the brain.

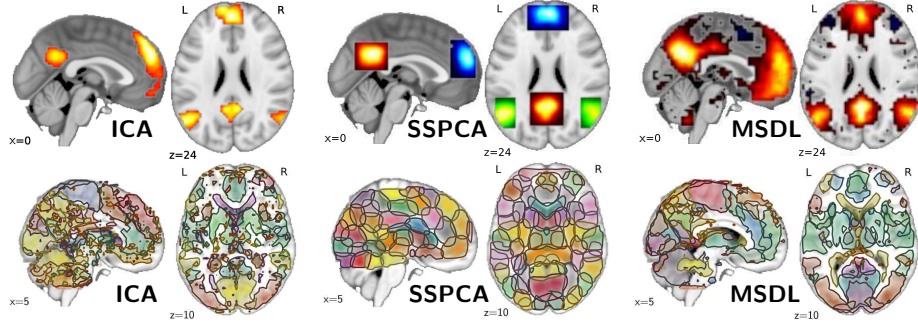


Fig. 1. Output of different decomposition models: from left to right, ICA, SSPCA and MSDL. **top:** default-mode network; for ICA maps are thresholded but for SSPCA and MSDL the transparency indicates zeros resulting from sparsity. **bottom:** outline of all maps. Some ICA-extracted maps show little structure and appear as noise. On the opposite, all the SSPCA maps segment regions, but their spatial regularity may be considered as forced. MSDL maps are more appealing: less noisy and less geometric.

2.3 Multi-subject analysis, goldmine or minefield?

To increase statistical power, data is often pooled across subjects. Yet inter-subject modeling of resting-state is challenging as the spatial maps differ across subjects and the time-series are not related. Most often single-subject procedures are applied to the time-concatenated volumes of the different subjects [7]. Such concatenation models the data in the different subjects as drawn from the same distribution, *i.e.* sharing the same spatial maps \mathbf{V} . However, for ICA-based approaches, separating signal-space selection at the subject and the group level stabilizes estimation in high-model-order settings and extracts more meaningful maps [18]. In a sparse decomposition framework, group-level maps can be formulated as a prior on subject-level models, as in the *Multi-Subject Dictionary Learning* (MSDL) approach [19]. The corresponding joint estimation of subject-level and group-level maps yields a reference population atlas of brain networks and corresponding subject-specific counter-parts. While multi-subject analyses do stabilize model estimation (see Fig. 1), at very high model order the spatial maps may reflect modes of inter-subject variability, such as spatial gradients of some of the networks [19].

3 Connectivity graphs: methods shape our understanding

3.1 Disentangling indirect effects to infer functional connections

Given brain regions, an important problem is to quantify their functional interactions from fMRI observations. Traditionally, an opposition is made between *functional connectivity* models, that describe the correlations in the time-series, and *effective connectivity* models, reflecting the influence neural systems exerts over each other. In our opinion, this dichotomy is over-emphasized and there is a

continuum of models ranging from descriptive statistics on correlation matrices to inference of directed graphical models. Unless the number of regions is very small, graphical model estimation is ill-posed in the small-sample limit as the likelihood of a connection between a pair of regions is a function of the signal observed in these regions *conditionally on all the other regions*. Indeed, isolating the signal specific to that pair is necessary to disambiguate direct links from indirect effects. The simplest probabilistic model for fMRI is the multivariate normal, for which estimating graphical models boils to partial-correlation estimation. As with effective connectivity, these graphical models depict coupling between brain regions, yet as they involve only second-order statistics, they are close to correlations analysis usually classified as functional connectivity [14].

3.2 Regularizing inverse covariance estimation

As partial correlations are given by the inverse covariance matrix, the sparsity pattern of this matrix captures the graph of conditional independence between brain regions and it is a good candidate for brain connectivity graphs [15]. If the number of observations is not very large compared to the number of regions, the estimation of the inverse covariance must be regularized. Different regularization schemes yield different priors for connectivity structures [20]. To fit to the data, ℓ_1 penalized estimators create overly dense graphs. Greedy methods based on the PC algorithm, that proceed by pruning the graph, fail in the presence of high-degree nodes [20]. Finally top-down approaches decomposing the graph in sub-graphs –known as decomposable models in Gaussian graphs– are limited by the absence of small separable sub-systems in the brain [20].

In our opinion, a promising approach for extracting brain connectivity relies on imposing a common graph structure across subjects to estimate jointly multiple subject-specific models [16]. A penalty based on mixed-norms can set such population-level sparsity in a well-behaved –convex– estimation problem. The resulting population prior accumulates samples for better structure estimation.

3.3 Focusing on graph modularity rather than small-worldness

A hallmark of many current brain connectivity studies is their focus on small-world properties of brain graphs: efficient transport with very few connections [6]. However, in empirically-defined networks, these properties are very fragile to noise [3, 20]. For instance a bad control on false detections of connections, as with non-regularized inverse covariance estimates, will create spurious random links in essence fabricating a small-world network. On the other hand, small-world connectivity is hard to recover [2] and estimation procedures fail on high-degree nodes or graphs with many large loops. Given these pitfalls, what evidence on brain connectivity can we gather from fMRI?

The failure of decomposable models and greedy approaches suggests that a good representation of brain connectivity displays a large graph bandwidth and contains hub nodes. The brain networks identified by spatial analysis suggest that brain connectivity contains large structures. An appealing picture is that

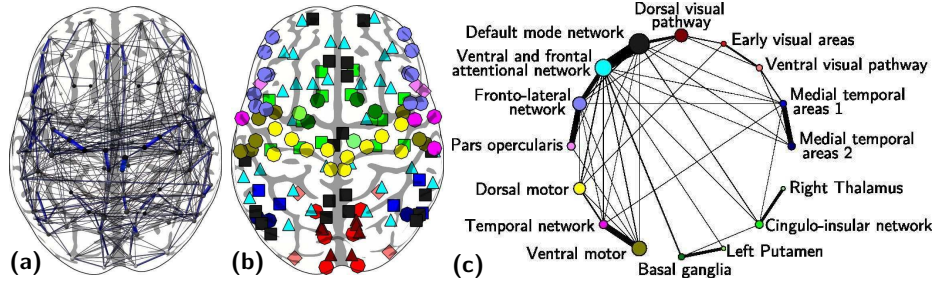


Fig. 2. (a) Functional connectivity learned with population-sparse inverse covariance; (b) corresponding graph communities; (c) inter-communities interactions [16].

of a modular graph composed of *communities* with inter-communities connections large in comparison with extra-communities connections [11, 6]. Identifying these communities is a clustering task. It is general a non-convex problem with little guarantees of success from the learning theory standpoint. It can be partly relaxed in a better-behaved graph-cut problem [16]. The resulting functional-segregation structure highlights from the connectivity graphs large clique-like sub-systems that match well the known brain networks (Fig. 2) [16]. The full-brain connectivity graph in which they are embedded also reveals their functional interactions (Fig. 2).

4 Conclusion: with great priors comes great responsibility

Probabilistic modeling of resting-state is a hard problem because it entails fitting very-high-dimensional models in unsupervised settings. Going beyond simple pictures requires moving away from descriptive statistics and applying state-of-the-art statistical learning techniques such as adequate penalizations on the various estimation steps. However penalizations are nothing but well-posed formulations of priors, and they can shape the output of the methods. These priors must not only be consistent with current neuroscientific knowledge but also robust to the confounding structured noise of fMRI.

Performing principled model selection to pick the best performing prior is still an open question, in the absence of clearly defined statistical task. Indeed, simply selecting for goodness of fit can inflate artificially model complexity to explain structured noise or intersubject variability [19]. For instance, it fails to capture an optimal scale for the description. In addition, while model fit on test data may guide the choice of a penalty, it is intrinsic to the model postulated.

Over-interpretation of plausible results is a recurrent temptation. Given the shortage of observations compared to the complexity of the models, how much can we conclude from the data, and how much is embedded in our priors? We believe that existing large multi-subject and multi-centric databases [5] can help to overcome these questions as they call for less regularization and can establish and calibrate priors for smaller studies.

References

1. Beckmann, C.F. Smith, S.M.: Probabilistic independent component analysis for functional magnetic resonance imaging. *Trans Med Im* 23, 137 (2004)
2. Bento, J. Montanari, A.: Which graphical models are difficult to learn? In: *Adv NIPS*, p. 1303 (2009)
3. Bialonski, S., Horstmann, M., Lehnertz, K.: From brain to earth and climate systems: Small-world interaction networks or not? *Chaos* 20, 013134 (2010)
4. Biswal, B., Zerrin Yetkin, F., Haughton, V., Hyde, J.: Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magn Reson Med* 34, 53719 (1995)
5. Biswal, B., et al.: Toward discovery science of human brain function. *Proc Ntl Acad Sci* 107, 4734 (2010)
6. Bullmore, E. Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10, 186 (2009)
7. Calhoun, V.D., Adali, T., Pearlson, G.D., Pekar, J.J.: A method for making group inferences from fMRI data using independent component analysis. *Hum Brain Mapp* 14, 140 (2001)
8. Cole, D., Smith, S., Beckmann, C.: Advances and pitfalls in the analysis and interpretation of resting-state fmri data. *Frontiers in systems neuroscience* 4 (2010)
9. Damoiseaux, J.S., et al.: Consistent resting-state networks across healthy subjects. *Proc Natl Acad Sci* 103, 13848 (2006)
10. Daubechies, I., et al.: Independent component analysis for brain fMRI does not select for independence. *Proc Natl Acad Sci* 106, 10415 (2009)
11. He, Y., et al.: Uncovering intrinsic modular organization of spontaneous brain activity in humans. *PLoS One* 4, e5226 (2009)
12. Jenatton, R., Obozinski, G., Bach, F.: Structured sparse principal component analysis. In: *Proc. AISTATS* (2010)
13. Kiviniemi, V., et al.: Functional segmentation of the brain cortex using high model order group PICA. *Hum brain map* (2009)
14. McIntosh, A.: Moving between functional and effective connectivity. In: Sporns, O. (ed.) *Analysis and Function of Large-Scale Brain Networks*, p. 15. Society for Neuroscience (2010)
15. Smith, S., et al.: Network modelling methods for fMRI. *Neuroimage* 54, 875 (2011)
16. Varoquaux, G., Gramfort, A., Poline, J.B., Thirion, B.: Brain covariance selection: better individual functional connectivity models using population prior. In: *Adv NIPS* (2010)
17. Varoquaux, G., Keller, M., Poline, J., Ciuciu, P., Thirion, B.: ICA-based sparse features recovery from fMRI datasets. In: *ISBI*. p. 1177 (2010)
18. Varoquaux, G., et al.: A group model for stable multi-subject ICA on fMRI datasets. *NeuroImage* 51, 288 (2010)
19. Varoquaux, G., Gramfort, A., Pedregosa, F., Michel, V., Thirion, B.: Multi-subject dictionary learning to segment an atlas of brain spontaneous activity. In: *Information Processing in Medical Imaging*. p. 562 (2011)
20. Varoquaux, G., Gramfort, A., Poline, J.B., Thirion, B.: Markov models for fMRI correlation structure: is brain functional connectivity small world, or decomposable into networks? *Journal of Physiology - Paris* p. epub ahead of print (2012)